REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
03-09-2015	Final	01/2013-06/2015

4. TITLE AND SUBTITLE

Enhancing listener strategies using a payoff matrix in speech-on-speech masking experiments

5a. CONTRACT NUMBER
FA8650-14-D-6501-0003

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Thompson, Eric R.
Iyer, Nandini
Simpson, Brian D.
Wakefield, Gregory H., Kieras, David E., Brungart, Douglas S.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
H0JS (2313CB14)

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
AND ADDRESS(ES)
Air Force Research Laboratory, 711 Human Performance Wing, Wright-Patterson AFB, Ohio 45433
Electrical Engineering & Computer Science, University of Michigan, Ann Arbor, MI 48109
Walter Reed National Military Medical Center, Bethesda, MD 20889

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Air Force Office Of Scientific Research
Arlington, VA 32034

10. SPONSOR/MONITOR'S ACRONYM(S)
AFOSR

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Distribution A: Approved for public release; distribution unlimited.

13. SUPPLEMENTARY NOTES
88ABW Cleared 11/21/2014; 88ABW-2014-5369.

14. ABSTRACT
Speech recognition was measured as a function of the target-to-masker ratio (TMR) with a syntactically similar speech masker. In the first experiment, the listeners were instructed to report the keywords from the target sentence. Data averaged across listeners showed a plateau in performance below 0 dB TMR when the masker and target sentences were from the same talker. The data showed that some listeners tended to report the target words at all TMRs in accordance with the instructions, while others reported keywords from the louder of the sentences, contrary to the instructions. In the second experiment, the stimuli were the same as in the first experiment, but the listeners were also instructed to avoid reporting the masker keywords, and a payoff matrix penalizing masker keywords and rewarding target keywords was used. In this experiment, the listeners reduced the number of reported masker keywords, and increased the number of reported target keywords overall, and the average data showed a local minimum at 0 dB TMR with same-talker maskers. The best overall performance with a same-talker masker was obtained with a level difference of 9 dB, where listeners achieved near perfect performance when the target was louder, and at least 80% correct performance when the target was the quieter of the two sentences.

15. SUBJECT TERMS
Speech intelligibility, speech masking, stream segregation, speech communication, listener strategy, auditory cognition

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			Eric Thompson
Unclassified	Unclassified	Unclassified		9	**19b. TELEPHONE NUMBER** *(include area code)*

Enhancing listener strategies using a payoff matrix in speech-on-speech masking experiments[a)]

Eric R. Thompson,[1,b)] Nandini Iyer,[1] Brian D. Simpson,[1] Gregory H. Wakefield,[2] David E. Kieras,[2] and Douglas S. Brungart[3]

[1]*Battlespace Acoustics Branch, Air Force Research Laboratory, 2610 Seventh Street B441, Wright-Patterson Air Force Base, Ohio 45433, USA*

[2]*Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, Michigan 48109, USA*

[3]*Walter Reed National Military Medical Center, 8901 Rockville Pike, Bethesda, Maryland 20889, USA*

(Received 7 November 2014; revised 3 June 2015; accepted 29 July 2015; published online 3 September 2015)

Speech recognition was measured as a function of the target-to-masker ratio (TMR) with syntactically similar speech maskers. In the first experiment, listeners were instructed to report keywords from the target sentence. Data averaged across listeners showed a plateau in performance below 0 dB TMR when masker and target sentences were from the same talker. In this experiment, some listeners tended to report the target words at all TMRs in accordance with the instructions, while others reported keywords from the louder of the sentences, contrary to the instructions. In the second experiment, stimuli were the same as in the first experiment, but listeners were also instructed to avoid reporting the masker keywords, and a payoff matrix penalizing masker keywords and rewarding target keywords was used. In this experiment, listeners reduced the number of reported masker keywords, and increased the number of reported target keywords overall, and the average data showed a local minimum at 0 dB TMR with same-talker maskers. The best overall performance with a same-talker masker was obtained with a level difference of 9 dB, where listeners achieved near perfect performance when the target was louder, and at least 80% correct performance when the target was the quieter of the two sentences. [http://dx.doi.org/10.1121/1.4928395]

[JFC] Pages: 1297–1304

I. INTRODUCTION

The ability to understand speech in the presence of speech maskers has been a subject of study at least since Miller (1947), who measured target speech recognition as a function of the number of masker talkers. Cherry (1953) suggested a number of cues that may aid a listener in segregating multiple simultaneous talkers such as differences in spatial location, voice quality, pitch, accent, or subject matter, and discussed the confusion that listeners experienced when trying to follow one utterance when a second utterance by the same talker was presented simultaneously. In more recent studies, researchers have investigated the effects of varying some of the cues suggested by Cherry (1953) for segregating competing talkers. A difference in voice pitch can be an effective cue for speech segregation, resulting in lower intelligibility for a male (target) talker with another male (masker) talker, than for the same male talker with a female masker talker (Festen and Plomp, 1990), and even lower intelligibility when the masker talker is the same as the target talker (with recorded or synthesized speech; Brungart, 2001; Cooke *et al.*, 2008).

Egan *et al.* (1954) and Dirks and Bower (1969) measured speech intelligibility with a single, same-talker competing speech masker, and reported a plateau in performance between −10 and 0 dB target-to-masker ratio (TMR), with performance increasing for TMRs greater than 0 dB, and decreasing for TMRs less than −10 dB. They suggested that this plateau appeared for conditions in which there were few segregation cues (i.e., with the same talker for both target and masker, and with only small intensity differences between target and masker). Brungart (2001) found evidence for a similar plateau in performance with same-talker maskers for negative TMRs between −12 and 0 dB, as well as evidence suggesting that performance could *increase* with decreasing TMRs below 0 dB. Brungart and Simpson (2007) and Cooke *et al.* (2008) both reported an increase in performance of about 10% points from 0 dB TMR to −9 dB TMR with same-talker maskers. This non-monotonicity in performance with TMR was also evident in data reported with a same-sex masker, but not with a different-sex masker (Brungart *et al.*, 2001) where the different pitches of the target and masker may have been sufficient for segregation at all TMRs. The listeners' ability to use a level cue to segregate a quieter target talker (in the absence of other cues, e.g., pitch differences) appears to only be effective with one interfering talker. When a second interfering talker is added (Brungart *et al.*, 2001; Iyer *et al.*, 2010), or when a noise masker is added (Agus *et al.*, 2009; Iyer *et al.*, 2010), performance degrades monotonically with decreasing TMR, and the data can be described by a smooth ogival psychometric function (cf. Eddins and Liu, 2012).

For a same-talker masker at 0 dB TMR, it is not surprising that listeners would have difficulty knowing which were

[a)]Portions of these data were presented at the 2014 International Conference on Auditory Display, New York, June 22–25.

[b)]Electronic mail: eric.thompson.28@us.af.mil

the target words and which were the masker words because, on average, both utterances will have the same level and approximately the same pitch. If both target and masker are drawn from a small, closed set of keywords, then it would be expected that listeners would have a lot of uncertainty about the target utterance, and there would be a lot of masker word reports. Indeed, both Brungart (2001) and Cooke *et al.* (2008) reported increased masker keywords at 0 dB TMR for same-talker masking conditions. At negative TMRs, the masker sentence should be intelligible, as is the case with the target at corresponding positive TMRs. Therefore, in a design where target and masker words are mutually exclusive [as was the case for Brungart (2001) and Cooke *et al.* (2008)], one might expect that if a listener hears intelligible masker words, but does not hear an intelligible target due to the adverse TMR, the listener should *never* respond with the masker words, but should eliminate the masker words from the response set, thereby making a more informed guess from the remaining words in the closed set. This optimal strategy would show a response pattern where the proportion of masker words reported would be at a maximum at 0 dB TMR, due to the absence of reliable pitch or level cues, but would decrease toward zero for both positive and negative TMRs. However, both Brungart (2001) and Cooke *et al.* (2008) reported that more than half of the errors at their lowest measured TMRs (−12 and −9 dB, respectively) were substitution errors, where masker words were reported instead of target or other words from the response set. Brungart and Simpson (2004) attempted to induce more optimal strategies in their listeners by providing feedback (no feedback was provided by Brungart, 2001), and also by reducing masker uncertainty by freezing the content of a masker throughout a block of trials in a three-talker context. They reported a small reduction in masker word reports with feedback, and that the listeners were unable to fully optimize their response strategies to eliminate the masker keywords from their responses even when masker uncertainty was reduced by repeating the identical masker waveform throughout the whole block of trials.

It appears that listeners can take advantage of level differences between two simultaneous same-talker utterances and even report words spoken by the quieter talker. However, it is not clear why the most common error is a substitution of masker words at the low TMRs. The current experiment was conducted to assess if listeners could use a more optimal response strategy that reduced the proportion of masker word reports at low TMRs through the use of a payoff matrix that provided incentives for reporting target words and penalties for reporting masker words.[1] Further, previous results are inconsistent regarding whether one should expect an increase in performance below 0 dB TMR (e.g., Brungart and Simpson, 2007; Cooke *et al.*, 2008) with same-talker maskers or a plateau (e.g., Egan *et al.*, 1954; Brungart, 2001). It is also unclear at what (negative) TMR one should expect performance to decrease again. The current study was conducted to address these issues by revisiting the intelligibility of speech in a two-talker situation with same-talker, same-sex and different-sex maskers.

II. EXPERIMENT 1

A. Listeners

Eighteen listeners (nine female) between the ages of 19 and 31 (median age 24) participated in the experiment. All had clinically normal hearing with audiometric thresholds of 20 dB hearing level (HL) or less in both ears at octave frequencies from 125 Hz to 8 kHz. They had prior experience with similar experiments, and they were paid for their participation.

B. Stimuli

The listeners heard a target sentence drawn randomly from the Coordinate Response Measure (CRM) corpus (Bolia *et al.*, 2000), which was also used by Brungart (2001). These sentences have the form "Ready, <call sign>, go to <color>-<number> now," where the call sign keyword is one of eight choices (Arrow, Baron, Charlie, Eagle, Hopper, Laker, Ringo, or Tiger), the color keyword is one of four choices (blue, green, red, or white), and the number keyword is one of the integers 1–8. The target sentence in the present study always had the call sign "Baron." The listeners also heard a masker sentence, also drawn from the CRM corpus and presented simultaneous to the target sentence. The sentences were selected randomly for each trial so that the target and masker sentences never had the same call sign, color, or number. The masker sentence was presented at 65 dB sound pressure level (SPL), and the target sentence level was adjusted according to the TMR selected randomly for each trial, ranging from −18 dB to +9 dB in 3-dB increments. On about 2% of the trials, the target was presented at 65 dB SPL with no masker.

C. Equipment

The stimuli were processed and the experiment was controlled in MATLAB (The Mathworks, Natick, MA, v2013a, 64-bit) running on a Windows 7 PC. Stimuli were presented to the listeners via an RME sound card (Hammerfall DSP Multiface II) over headphones (Sennheiser HD 280 PRO) in a sound-attenuating listening booth. Instructions and a response GUI were presented to the listeners on a computer monitor in the booth, and responses were made using a mouse by clicking on the appropriate button on an 8×4 grid of colored and numbered buttons on the GUI corresponding to the CRM corpus.

D. Procedure

The listeners were instructed to report the color-number pair from the sentence addressed to the call sign Baron. They completed blocks of 51 trials, within which the masker configuration (same talker, same sex, or different sex, with respect to the target talker) was fixed. Within each block of trials, the TMR was pseudorandomly selected for each trial so that each listener performed 40 trials in total for each TMR and masker configuration. There were also 24 trials in which the target was presented without a masker, randomly interspersed within the blocks. These unmasked trials were

included to obtain an estimate of the lapse rate for the psychometric function. Therefore, each listener completed 1224 trials (40 trials × 10 TMRs × 3 masker configurations + 24 unmasked trials). The response GUI was visible on the screen throughout the block of trials, but was only enabled for input after the end of the stimulus. After the listener selected a response, the GUI would indicate the correct response by flashing the background on the button corresponding to the target color-number pair for about 1 s. After the feedback period, the experiment would automatically continue on to the next trial.

E. Results

The breakdowns of response proportions, averaged across listeners, are shown in Fig. 1 as a function of TMR, with the data from the current study plotted with open symbols along with the data from Brungart (2001) replotted here with gray-filled symbols. The data are shown here separated by color and number responses because the patterns of proportion of correct responses was different for the two keywords in the Brungart (2001) data. As can be seen in Fig. 1, the data from the current study are similar to those from the 2001 data collection (gray-filled and open symbols in Fig. 1, respectively), increasing from about 55% target word (color and number) responses at 0 dB TMR to more than 90% target word responses at 9 dB TMR in the same-talker condition (left panels) for both color and number responses (circle markers in Fig. 1). The subjects reported the correct target words on greater than 99% of the trials when no masker was presented. The same-talker, number response data from both data sets show an increase in performance with TMRs decreasing from 0 dB to a maximum of about 75% correct between −9 and −12 dB TMR. The only divergence between the two data sets is for the color responses at negative TMRs. In the 2001 data (gray-filled symbols in Fig. 1), the proportion of target color responses is roughly independent of TMR for TMRs less than 0 dB with a same-talker masker, indicating the plateau reported by Egan *et al*. (1954) and Dirks and Bower (1969), with the target color being reported about 48% of the time. However, in the current data (open symbols in Fig. 1), the proportion of target color responses increases with decreasing TMR to a maximum of about 63% at −9 dB TMR. In all of the data shown in Fig. 1, almost all of the responses were either target words (circles) or masker words (squares). The other word responses (i.e., neither target nor masker words) comprise fewer than 5% of all responses at all TMRs except the lowest measured TMRs.

F. Discussion

The same-talker condition is challenging for the listeners, especially around 0 dB TMR, because there are minimal pitch or level cues to differentiate the two sentences. While both target and masker sentences may be intelligible, they are very confusable, and the listeners may not be able to keep track of which words were from the sentence addressed to Baron. Therefore, it is not surprising that the proportions of target and masker words are each close to 50% for colors

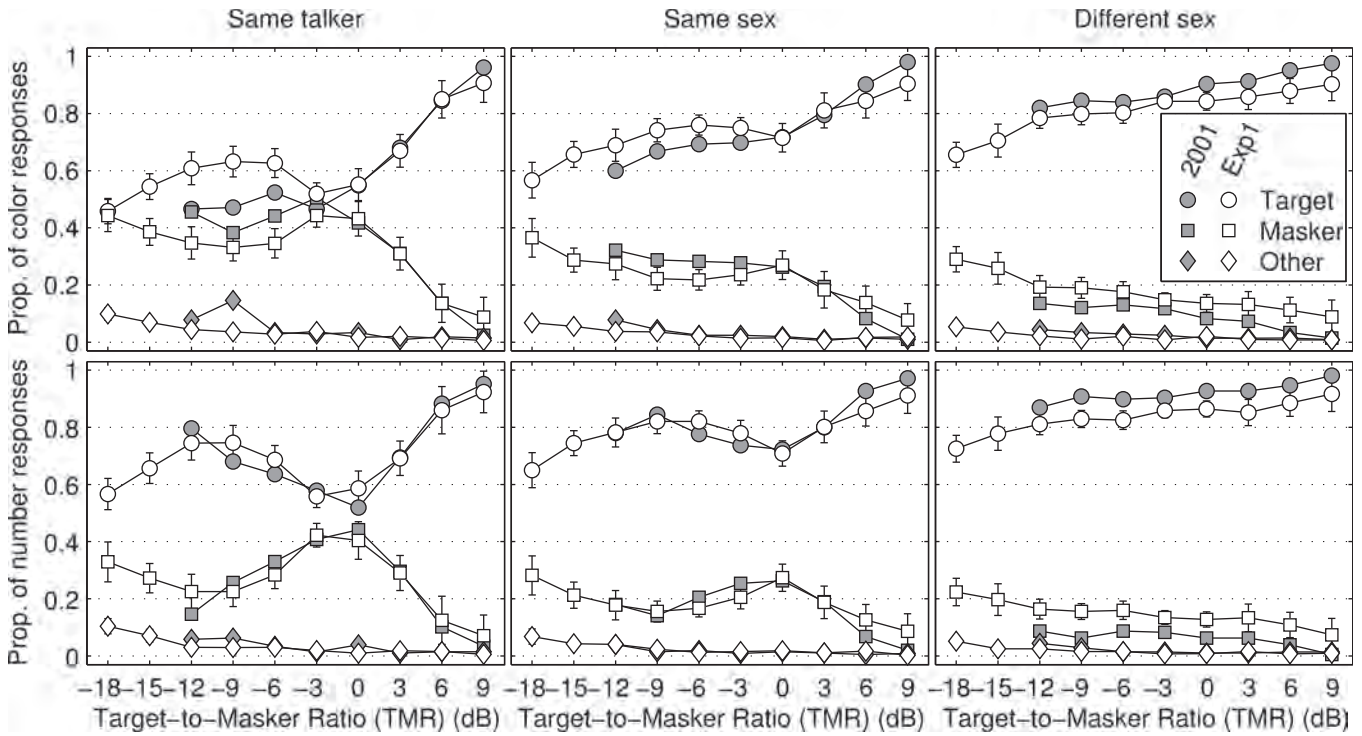

FIG. 1. Proportion of color responses (top panels) and number responses (bottom panels) for target (circles), masker (squares), and other (diamonds) responses from experiment 1 as a function of TMR. The error bars in Fig. 1 represent 95% confidence intervals for within-subject designs (Morey, 2008, based on Loftus and Masson, 1994). The left panels show data obtained with same-talker maskers, the center panels with same-sex maskers, and the right panels with different-sex maskers. The data from the present study are plotted with open symbols, and the data from Brungart (2001) are replotted here with gray-filled symbols (with permission from the author).

and numbers at 0 dB TMR (see Fig. 1). As the TMR increases from 0 dB, the target energy dominates the masker, so the target words should become more intelligible, and the listeners should be able to use the higher relative sentence level as a cue for identifying the words in the target sentence. Conversely, as the TMR decreases from 0 dB, the masker energy dominates the target, so the listeners would have to use the lower relative sentence level to determine which words belong to the target sentence. As long as the target sentence is still audible at the negative TMRs, the listeners should be able to improve their performance over the 0 dB TMR conditions by reporting the color-number pair from the quieter sentence. However, when the TMR is negative, it may be difficult for the listeners to understand the words from the target sentence because the energy of the masker sentence is dominant. In this case, the masker sentence should be intelligible and, assuming that the listeners know that the masker and target words are drawn from a closed set without replacement, the listeners should be able to make a more informed guess by eliminating the masker words from the response set. This strategy would be seen in the data as a decrease in masker word responses with decreasing TMR below 0 dB TMR in favor of a higher proportion of other (i.e., not target or masker) word reports. This pattern was not observed in the data from the current study or from Brungart (2001).

Investigations into individual performance revealed that some of the listeners had a greater tendency to report the masker than other listeners. The data from the listener with the most masker word responses, and from the listener with the fewest masker word responses are shown in Fig. 2 with the open symbols indicating performance for the listener with the fewest masker responses (S_1), and the filled symbols indicating performance for the listener with the most masker responses (S_2). The data from the two listeners are quite similar at positive TMRs for both colors and numbers. With negative TMRs, the two listeners' data curves again look quite similar at first glance, especially for color responses, except that while one listener is reporting the target words (circles), the other listener is reporting the masker words (squares). It appears from these data that these two listeners were using very different task strategies: S_1 was following the instructions and trying to report the color-number pair addressed to Baron, and S_2 tended to use a "use what you heard" strategy, reporting the color-number from the louder sentence in the same-talker masking conditions. The plateau observed in the data from previous studies with negative TMRs for same-talker maskers may just be the result of a combination of these two strategies. The data from three listeners in experiment 1 clearly showed the pattern exemplified by S_2's data in Fig. 2 (including S_2), and six listeners' data clearly showed S_1's pattern (including S_1). The other subjects' data showed a pattern between these two extremes that was more similar to the overall average. On account of these differences in apparent task strategy, experiment 2 was designed to encourage all of the listeners to use a more homogeneous strategy through more explicit instructions and a payoff matrix, which rewarded correct (target) responses, and penalized masker responses.

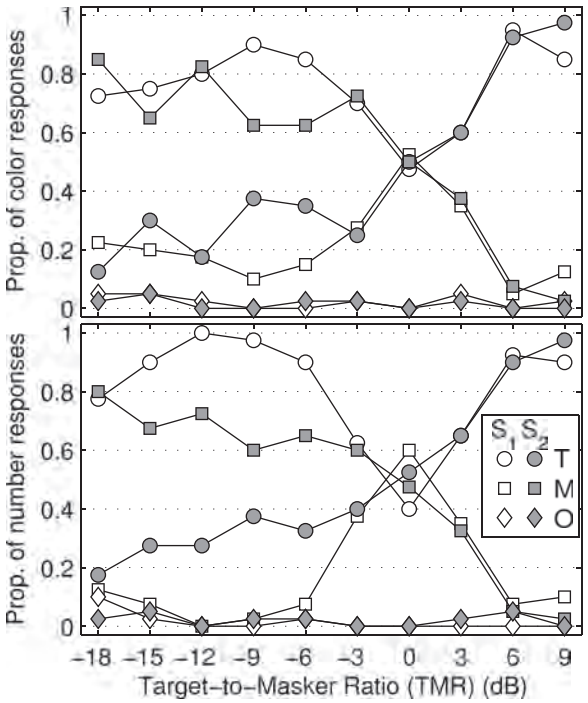

FIG. 2. Proportion of color (top panel) and number responses (bottom panel) for target (T, circles), masker (M, squares), and other (O, diamonds) responses for the same-talker condition for the listener who reported the fewest (S1, open symbols) and the most (S2, gray-filled symbols) masker words from experiment 1 as a function of TMR.

In the Brungart (2001) study, data were collected in blocks of trials without feedback, and with randomly selected masker talkers for every trial (i.e., the same-talker, same-sex, and different-sex masking conditions were combined together randomly within a block of trials). In the present study, the masking condition was fixed throughout each block of trials (e.g., always same-talker masker), and feedback was provided to the listeners after every response, informing them what the correct response was. Despite these procedural differences, the data turned out to be quite similar across the two studies, which suggests that these two factors, masking condition consistency within a block and feedback, did not have a large impact on performance.

III. EXPERIMENT 2

The stimuli and equipment used in experiment 2 were identical to those used in experiment 1. The only differences between experiments 1 and 2 were a slightly different pool of listeners, different instructions, and the introduction of a payoff matrix designed to reward desired behavior, as detailed below.

A. Listeners

Nine of the listeners from experiment 1 (four female) also participated in experiment 2. The data from experiment 1 for three of these listeners showed the trend exemplified by S_1 in Fig. 2, and for another three showed the trend exemplified by S_2. One additional female listener, who had not participated in experiment 1, was added to the listener panel for

experiment 2, for a total of ten listeners. She also had clinically normal hearing of 20 dB HL or less in both ears at octave frequencies from 125 Hz to 8 kHz. The age range for this experiment was 19 to 32 years (median age 24 years).

B. Procedure

The procedure used in experiment 2 was similar to that used in experiment 1, but was changed only in terms of the instructions and feedback provided to the listeners. Before the experiment, the listeners were verbally instructed that their task was to report the color-number pair addressed to the Baron call sign, as in experiment 1. They were also explicitly instructed not to report the color-number pair addressed to any call sign other than Baron, and that if they only heard one sentence and it was not addressed to Baron they should guess a color and number that were not in that sentence. These instructions were also repeated in writing on the GUI at the beginning of every block of trials. They were also informed that they would receive 1 point for every correct target word (color and number) response, and that they would *lose* 2 points for every reported masker word (color and number) response, and no points would be gained or lost for reporting any words other than the target or masker. If they responded correctly with both target words on every trial, the maximum number of points achievable was 2400, whereas if they responded with both masker words on every trial, the minimum possible points was −4800. In addition, they would receive a bonus (equivalent to two hours of time as a subject) for achieving a total of 1200 points at the end of the experiment. At the end of every block, they were shown a plot of the cumulative sum of their points up to and including their current block as a function of the trial number across the blocks, along with the cumulative sums of all of the other (anonymous) listeners. A straight line connecting (0,0) with (1200,1200) was also displayed to indicate the trend required to achieve the goal over the 1200 trials of the experiment (not counting the 24 catch trials).

C. Results

Comparing the same-talker data from experiments 1 and 2 (left panels of Figs. 1 and 3, respectively) shows that the listeners not only improved their correct response performance from experiment 1 to experiment 2, but also decreased the proportion of masker responses. The data are very similar between the two experiments for TMRs greater than or equal to 0 dB, with a minimum in the correct responses at 0 dB TMR of 52% for colors and 56% for numbers, and an increase to about 95% target word reports for +9 dB TMR. As with experiment 1, the subjects responded with the target words on more than 99% of all trials presented without a masker. The greatest improvement in performance between experiments 1 and 2 for the same-talker condition is observed for the negative TMRs, where the local maximum in target word reports at around −9 dB TMR increased from 63% to 78% for color responses, and from 75% to 88% for number responses, and the proportion of masker word responses (also at −9 dB TMR) decreased from 33% to 17% for colors, and from 22% to 8% for numbers. Performance

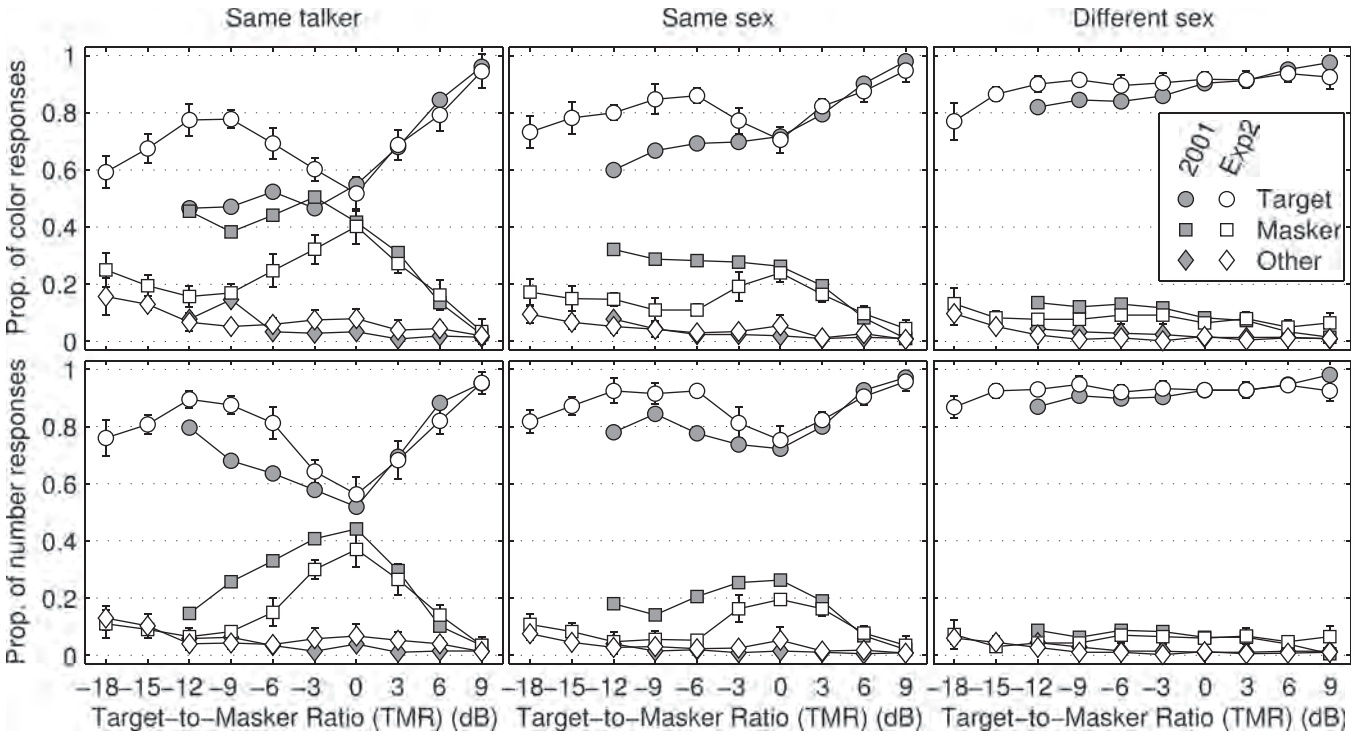

FIG. 3. Proportion of color (top panels) and number responses (bottom panels) for target (circles), masker (squares), and other (diamonds) responses from experiment 2 as a function of TMR. The error bars represent 95% confidence intervals. The left panels show data obtained with same-talker maskers, the center panels with same-sex maskers, and the right panels with different-sex maskers. The data from the present study are plotted with open symbols, and the data from Brungart (2001) are replotted here with gray-filled symbols (with permission from the author).

then drops off for TMRs less than −12 dB. The proportion of other (i.e., not target or masker) words remained very low, similar to experiment 1. A repeated-measures analysis of variance performed separately for color and number responses, with within-listeners factors of experiment (1 and 2) and TMR, using the same-talker masker data (rationalized arcsine transformed percent correct colors and numbers, Studebaker, 1985) from the nine listeners who participated in both experiments 1 and 2 showed a significant main effect of experiment [$F(1, 8) = 12.4$, $p < 0.01$ for colors, $F(1, 8) = 7.1$, $p < 0.05$ for numbers], a significant main effect of TMR [$F(9, 72) = 23.0$, $p < 0.001$ for colors, $F(9, 72) = 14.87$, $p < 0.001$ for numbers], and a significant interaction [$F(9, 72) = 6.8$, $p < 0.001$ for colors, $F(9, 72) = 4.7$, $p < 0.001$ for numbers]. A similar analysis was conducted on the proportion of masker responses (also rationalized arcsine transformed proportions), with similar results: significant main effect of experiment [$F(1, 8) = 23.82$, $p < 0.01$ for colors, $F(1, 8) = 15.88$, $p < 0.01$ for numbers], significant main effect of TMR [$F(9, 72) = 19.36$, $p < 0.001$ for colors, $F(9, 72) = 12.74$, $p < 0.001$ for numbers], and a significant interaction between TMR and experiment [$F(9, 72) = 6.3$, $p < 0.001$ for colors, $F(9, 72) = 4.9$, $p < 0.001$ for numbers].

The data from the same-sex and different-sex masker conditions from experiment 2 are shown in Fig. 3, center and right panels, respectively (top row: color responses; bottom row: number responses). In both the same-sex and different-sex masking conditions, performance was similar between experiments 1 and 2, although the listeners achieved slightly better performance at negative TMRs in experiment 2 than in experiment 1, along with a lower proportion of masker word reports. The dip in performance around 0 dB TMR is more pronounced in the same-sex masker data from experiment 2 than from experiment 1. There was about a 20%-point improvement in scores at 0 dB TMR from the same-talker to the same-sex masker condition, and another 20%-point improvement from the same-sex to the different-sex masker. With a TMR of −9 dB, these differences in scores between conditions were reduced to about 5% points.

The point scheme used in experiment 2 was applied in retrospect to the data from experiment 1 to compare overall performance across the two data sets. In experiment 1, where the listeners had not been explicitly instructed to avoid reporting masker words, the listeners' scores ranged from −809 to 1623, and only five of the eighteen listeners would have achieved the 1200 point goal, had it been in place for experiment 1. Seven out of the ten listeners in experiment 2, when they were explicitly instructed to avoid reporting masker words, achieved the 1200 point goal, with a range of scores for all listeners from 1103 to 1943. For the nine listeners who participated in both experiments, the mean increase in point total was 876 points, with a range of −108 to 2054 points of improvement (one listener had a decrease in score from 1519 to 1411 between experiments 1 and 2). This mean increase of 876 points was achieved through an average increase of 259 target word responses over the 2400 total responses (combining color and number responses), and an average decrease of 309 masker word responses (618 points).

In the discussion of experiment 1, the data from two listeners (fewest masker word reports and most masker word reports) were highlighted in Fig. 2. The data from experiment 2 from the same two listeners are shown in Fig. 4. The proportion of target words reported by the listener who had the fewest masker word responses in experiment 1 (S_1, open symbols in Figs. 2 and 4) stayed about the same in experiment 2, while the proportion of masker responses decreased dramatically between the two experiments, particularly around 0 dB TMR. The listener with the most masker word responses in experiment 1 (S_2, filled symbols), who had more masker responses for negative TMRs and more target responses for positive TMRs, reported more target than masker words at all TMRs in experiment 2 except around 0 dB TMR, where the responses were almost evenly split between target and masker. This listener seems to have switched strategies from tending to report the louder color-number pair in experiment 1 to listening for the color-number pair in the sentence addressed to Baron.

D. Discussion

With more specific instructions and incentives to promote the desired behavior of reporting target words while minimizing masker word reports, the listeners in this study were able to not only reduce the number of masker reports, but also to increase the number of target reports. It is assumed that if the listeners were only improving performance with more practice from experiment 1 to experiment 2, then overall performance might have improved while

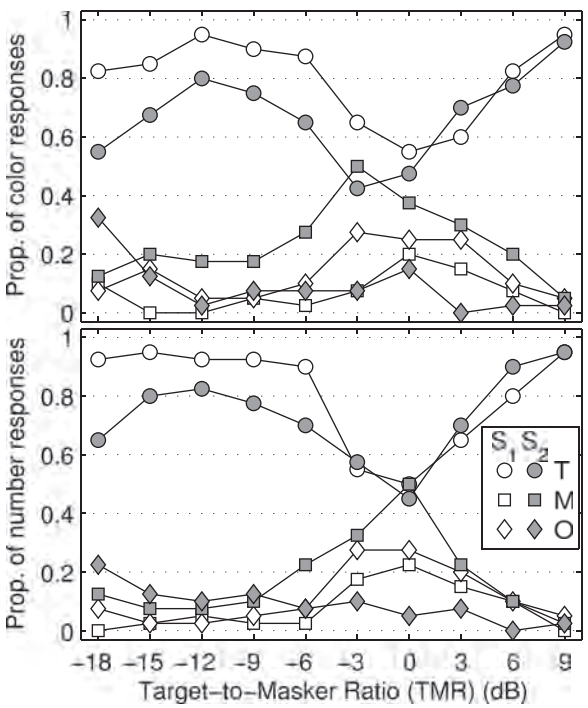

FIG. 4. Proportion of color (top panel) and number responses (bottom panel) for target (T, circles), masker (M, squares), and other (O, diamonds) responses for the same-talker condition for the listener who reported the fewest (S1, open symbols) and the most (S2, gray-filled symbols) masker words from experiment 2 as a function of TMR.

maintaining similar response patterns with a large proportion of masker word responses. However, the data show that the listeners were able to adopt a more appropriate (nearly optimal) response strategy by reducing the proportion of masker words reported at TMRs when the masker should have been clearly audible. Therefore, the results from experiment 2 resolve the discrepancy observed for color and number in the original experiment (Brungart, 2001), where a plateau was observed in the proportion of target color responses below 0 dB TMR, while the proportion of target number responses increased with decreasing TMRs below 0 dB. It is unknown how long these changes in behavioral patterns would last, and whether the listeners would revert back to reporting more masker words at negative TMRs in the absence of specific instructions and incentives.

The listeners could have further optimized their response strategy to increase their point total by increasing the proportion of other-word responses around 0 dB TMR. Assuming that the masker and target streams are both audible and intelligible around 0 dB TMR, or assuming that only the masker or the target was intelligible, but that the listeners were unsure whether they heard the target or the masker, the best strategy given the payoff matrix would be to respond with neither masker nor target, since the penalty for a masker word was greater than the gain for a target word. For the most points, the listeners should have only responded with what they heard when they thought it was twice as likely that what they heard was a target word than that it was a masker word. This strategy would have produced a pattern of higher proportions of other-word responses around 0 dB TMR. Since this pattern did not appear in the data, the listeners either did not come up with this strategy during the experiment, or they were relatively certain that what they were reporting was the target (even when incorrect). Adjusting the payoff matrix to more extreme values (e.g., lose 4 points for every masker word and gain 1 point for each target word) and repeating the experiment could disambiguate these two possibilities.

With both the same-talker masker and the same-sex masker, the performance curves showed a notch that was roughly symmetric and centered around 0 dB TMR. This notch reflects the degree of similarity between the target and masker utterances. When there was a large pitch difference between target and masker (as in the different-sex masker condition), there was no decrement in performance around 0 dB TMR because the pitch difference was sufficient to disambiguate target from masker. With the same-sex masker, the pitch differences between target and masker were much smaller. With a level difference greater than 6 dB, there was little difference in performance between the same-sex and different-sex masker conditions, which suggests that the listeners were able to use the level differences to select the target. This could reflect the expected variance of natural level fluctuations within each utterance, since level was equalized on a sentence, not individual word, basis. Level differences were likely to be even more important for the same-talker masker condition, since the pitch differences between utterances for the same talker would be even smaller than those between different, but same-sex maskers.

IV. CONCLUSIONS

In response to a payoff matrix, designed to reward target word responses and to penalize masker word responses, the listeners refined their strategies so that the number of masker word responses was reduced between experiments 1 and 2 in favor of more target responses. The data show that listeners are able to use level differences between two simultaneous talkers as a cue to select words from one talker or the other, even when the target talker is the (much) quieter of the two. With a level difference between the two talkers of about 9 dB, the listeners could identify nearly all of the target words from the louder of the two talkers, and about 80% of the target words when asked to report words from the quieter of the two talkers. This performance was achieved with all tested masking conditions (same-talker, same-sex, different-sex). This suggests that if one is listening to two simultaneous speech streams (with no other simultaneous sounds), the best configuration would be to adjust the level of one stream to be about 9 dB higher than the other. This configuration is especially important for situations in which one is listening to two highly confusable speech streams (similar pitch, similar content), where the only reliable segregation cue may be the level difference. If other segregation cues are available, introducing a level difference does not appear to have an adverse affect on performance, as evidenced by the performance in the different-sex masker conditions. However, if there are more than two speech streams, or if there is a significant noise source in addition to the two speech streams, then listeners may not be able to listen to the quietest talker (see Iyer *et al.*, 2010), and other strategies for enhancing speech segregation may need to be developed.

ACKNOWLEDGMENTS

This research was supported by grants from the Air Force Office of Sponsored Research (AFOSR) (N.I. and B.D.S.) and the Office of Naval Research (ONR) (Grant No. N00014-13-1-0358) (G.H.W. and D.E.K.).

[1]Payoff matrices have been used as a means of manipulating human behavior in many psychological studies, for example, related to signal detection theory (e.g., Green and Swets, 1988, pp. 21–23) and to game theory (e.g., Gray, 2002, pp. 560–561).

Agus, T. R., Akeroyd, M. A., Gatehouse, S., and Warden, D. (**2009**). "Informational masking in young and elderly listeners for speech masked by simultaneous speech and noise," J. Acoust. Soc. Am. **126**, 1926–1940.

Bolia, R. S., Nelson, W. T., Ericson, M. A., and Simpson, B. D. (**2000**). "A speech corpus for multitalker communications research," J. Acoust. Soc. Am. **107**, 1065–1066.

Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.

Brungart, D. S., and Simpson, B. D. (**2004**). "Within-ear and across-ear interference in a dichotic cocktail party listening task: Effects of masker uncertainty," J. Acoust. Soc. Am. **115**, 301–310.

Brungart, D. S., and Simpson, B. D. (**2007**). "Effect of target-masker similarity on across-ear interference in a dichotic cocktail-party listening task," J. Acoust. Soc. Am. **122**, 1724–1734.

Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (**2001**). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc. Am. **25**, 975–979.

Cooke, M., Garcia Lecumberri, M. L., and Barker, J. (**2008**). "The foreign language cocktail party problem: Energetic and informational masking effects in non-native speech perception," J. Acoust. Soc. Am. **123**, 414–427.
Dirks, D. D., and Bower, D. R. (**1969**). "Masking effects of speech competing messages," J. Speech Hear. Res. **12**, 229–245.
Eddins, D. A., and Liu, C. (**2012**). "Psychometric properties of the coordinate response measure corpus with various types of background interference," J. Acoust. Soc. Am. **131**, EL177–EL183.
Egan, J. P., Carterette, E. C., and Thwing, E. J. (**1954**). "Some factors affecting multi-channel listening," J. Acoust. Soc. Am. **26**, 774–782.
Festen, J. M., and Plomp, R. (**1990**). "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," J. Acoust. Soc. Am. **88**, 1725–1736.
Gray, P. (**2002**). *Psychology*, 4th ed. (Worth Publishers, New York), pp. 560–561.
Green, D. M., and Swets, J. A. (**1988**). *Signal Detection Theory and Psychophysics* (Peninsula Publishing, Los Altos, CA), pp. 21–23.
Iyer, N., Brungart, D. S., and Simpson, B. D. (**2010**). "Effects of target-masker contextual similarity on the multimasker penalty in a three-talker diotic listening task," J. Acoust. Soc. Am. **128**, 2998–3010.
Loftus, G. R., and Masson, M. E. J. (**1994**). "Using confidence intervals in within-subject designs," Psychon. Bull. Rev. **1**, 476–490.
Miller, G. A. (**1947**). "The masking of speech," Psychol. Bull. **44**, 105–129.
Morey, R. D. (**2008**). "Confidence intervals from normalized data: A correction to Cousineau (2005)," Tutor. Quant. Meth. Psychol. **4**, 61–64.
Studebaker, G. A. (**1985**). "A 'rationalized' arcsine transform," J. Speech Hear. Res. **28**, 455–462.